

**THE QUANTITY THEORY OF MONEY-A RESTATEMENT**

**Milton Friedman**

**Studies in the Quantity Theory of Money  
University Chicago Press  
1956**

## *The Quantity Theory of Money—A Restatement*

THE quantity theory of money is a term evocative of a general approach rather than a label for a well-defined theory. The exact content of the approach varies from a truism defining the term "velocity" to an allegedly rigid and unchanging ratio between the quantity of money—defined in one way or another—and the price level—also defined in one way or another. Whatever its precise meaning, it is clear that the general approach fell into disrepute after the crash of 1929 and the subsequent Great Depression and only recently has been slowly re-emerging into professional respectability.

The present volume is partly a symptom of this re-emergence and partly a continuance of an aberrant tradition. Chicago was one of the few academic centers at which the quantity theory continued to be a central and vigorous part of the oral tradition throughout the 1930's and 1940's, where students continued to study monetary theory and to write theses on monetary problems. The quantity theory that retained this role differed sharply from the atrophied and rigid caricature that is so frequently described by the proponents of the new income-expenditure approach—and with some justice, to judge by much of the literature on policy that was spawned by quantity theorists. At Chicago, Henry Simons and Lloyd Mints directly, Frank Knight and Jacob Viner at one remove, taught and developed a more subtle and relevant version, one in which the quantity theory was connected and integrated with general price theory and became a flexible and sensitive tool for interpreting movements in aggregate economic activity and for developing relevant policy prescriptions.

To the best of my knowledge, no systematic statement of this theory as developed at Chicago exists, though much can be read between the lines of Simons' and Mints's writings. And this is as it should be, for the Chicago tradition was not a rigid system, an unchangeable orthodoxy, but a way of looking at things. It was a theoretical approach that insisted that money does matter—that any interpretation of short-term movements in economic activity is likely to be seriously at fault if it neglects monetary changes and repercussions and if it leaves unexplained why people are willing to hold the particular nominal quantity of money in existence.

The purpose of this introduction is not to enshrine—or, should I say, inter—a definitive version of the Chicago tradition. To suppose that one

could do so would be inconsistent with that tradition itself. The purpose is rather to set down a particular "model" of a quantity theory in an attempt to convey the flavor of the oral tradition which nurtured the remaining essays in this volume. In consonance with this purpose, I shall not attempt to be exhaustive or to give a full justification for every assertion.

1. The quantity theory is in the first instance a theory of the *demand* for money. It is not a theory of output, or of money income, or of the price level. Any statement about these variables requires combining the quantity theory with some specifications about the conditions of supply of money and perhaps about other variables as well.

2. To the ultimate wealth-owning units in the economy, money is one kind of asset, one way of holding wealth. To the productive enterprise, money is a capital good, a source of productive services that are combined with other productive services to yield the products that the enterprise sells. Thus the theory of the demand for money is a special topic in the theory of capital; as such, it has the rather unusual feature of combining a piece from each side of the capital market, the supply of capital (points 3 through 8 that follow), and the demand for capital (points 9 through 12).

3. The analysis of the demand for money on the part of the ultimate wealth-owning units in the society can be made formally identical with that of the demand for a consumption service. As in the usual theory of consumer choice, the demand for money (or any other particular asset) depends on three major sets of factors: (a) the total wealth to be held in various forms—the analogue of the budget restraint; (b) the price of and return on this form of wealth and alternative forms; and (c) the tastes and preferences of the wealth-owning units. The substantive differences from the analysis of the demand for a consumption service are the necessity of taking account of intertemporal rates of substitution in (b) and (c) and of casting the budget restraint in terms of wealth.

4. From the broadest and most general point of view, total wealth includes all sources of "income" or consumable services. One such source is the productive capacity of human beings, and accordingly this is one form in which wealth can be held. From this point of view, "the" rate of interest expresses the relation between the stock which is wealth and the flow which is income, so if  $Y$  be the total flow of income, and  $r$ , "the" interest rate, total wealth is

$$W = \frac{Y}{r}. \quad (1)$$

Income in this broadest sense should not be identified with income as it is ordinarily measured. The latter is generally a "gross" stream with respect

to human beings, since no deduction is made for the expense of maintaining human productive capacity intact; in addition, it is affected by transitory elements that make it depart more or less widely from the theoretical concept of the stable level of consumption of services that could be maintained indefinitely.

5. Wealth can be held in numerous forms, and the ultimate wealth-owning unit is to be regarded as dividing his wealth among them (point [a] of 3), so as to maximize "utility" (point [c] of 3), subject to whatever restrictions affect the possibility of converting one form of wealth into another (point [b] of 3). As usual, this implies that he will seek an apportionment of his wealth such that the rate at which he *can* substitute one form of wealth for another is equal to the rate at which he is just willing to do so. But this general proposition has some special features in the present instance because of the necessity of considering flows as well as stocks. We can suppose all wealth (except wealth in the form of the productive capacity of human beings) to be expressed in terms of monetary units at the prices of the point of time in question. The rate at which one form can be substituted for another is then simply \$1.00 worth for \$1.00 worth, regardless of the forms involved. But this is clearly not a complete description, because the holding of one form of wealth instead of another involves a difference in the composition of the income stream, and it is essentially these differences that are fundamental to the "utility" of a particular structure of wealth. In consequence, to describe fully the alternative combinations of forms of wealth that are available to an individual, we must take account not only of their market prices—which except for human wealth can be done simply by expressing them in units worth \$1.00—but also of the form and size of the income streams they yield.

It will suffice to bring out the major issues that these considerations raise to consider five different forms in which wealth can be held: (i) money ( $M$ ), interpreted as claims or commodity units that are generally accepted in payment of debts at a fixed nominal value; (ii) bonds ( $B$ ), interpreted as claims to time streams of payments that are fixed in nominal units; (iii) equities ( $E$ ), interpreted as claims to stated pro-rata shares of the returns of enterprises; (iv) physical non-human goods ( $G$ ); and (v) human capital ( $H$ ). Consider now the yield of each.

(i) Money may yield a return in the form of money, for example, interest on demand deposits. It will simplify matters, however, and entail no essential loss of generality, to suppose that money yields its return solely in kind, in the usual form of convenience, security, etc. The magnitude of this return in "real" terms per nominal unit of money clearly

depends on the volume of goods that unit corresponds to, or on the general price level, which we may designate by  $P$ . Since we have decided to take \$1.00 worth as the unit for each form of wealth, this will be equally true for other forms of wealth as well, so  $P$  is a variable affecting the "real" yield of each.

(ii) If we take the "standard" bond to be a claim to a perpetual income stream of constant nominal amount, then the return to a holder of the bond can take two forms: one, the annual sum he receives—the "coupon"; the other, any change in the price of the bond over time, a return which may of course be positive or negative. If the price is expected to remain constant, then \$1.00 worth of a bond yields  $r_b$  per year, where  $r_b$  is simply the "coupon" sum divided by the market price of the bond, so  $1/r_b$  is the price of a bond promising to pay \$1.00 per year. We shall call  $r_b$  the market bond interest rate. If the price is expected to change, then the yield cannot be calculated so simply, since it must take account of the return in the form of expected appreciation or depreciation of the bond, and it cannot, like  $r_b$ , be calculated directly from market prices (so long, at least, as the "standard" bond is the only one traded in).

The nominal income stream purchased for \$1.00 at time zero then consists of

$$r_b(0) + r_b(0) d \left( \frac{1}{r_b(t)} \right) = r_b(0) - \frac{r_b(0)}{r_b^2(t)} \frac{d r_b(t)}{dt}, \quad (2)$$

where  $t$  stands for time. For simplicity, we can approximate this functional by its value at time zero, which is

$$r_b - \frac{1}{r_b} \frac{d r_b}{dt}. \quad (3)$$

This sum, together with  $P$  already introduced, defines the real return from holding \$1.00 of wealth in the form of bonds.

(iii) Analogously to our treatment of bonds, we may take the "standard" unit of equity to be a claim to a perpetual income stream of constant "real" amount; that is, to be a standard bond with a purchasing-power escalator clause, so that it promises a perpetual income stream equal in nominal units to a constant number times a price index, which we may, for convenience, take to be the same price index  $P$  introduced in (i). The nominal return to the holder of the equity can then be regarded as taking three forms: the constant nominal amount he would receive per year in

1. This is an oversimplification, because it neglects "leverage" and therefore supposes that any monetary liabilities of an enterprise are balanced by monetary assets.

the absence of any change in  $P$ ; the increment or decrement to this nominal amount to adjust for changes in  $P$ ; and any change in the nominal price of the equity over time, which may of course arise from changes either in interest rates or in price levels. Let  $r_e$  be the market interest rate on equities defined analogously to  $r_b$ , namely, as the ratio of the "coupon" sum at any time (the first two items above) to the price of the equity, so  $1/r_e$  is the price of an equity promising to pay \$1.00 per year if the price level does not change, or to pay

$$\frac{P(t)}{P(0)} \cdot 1$$

if the price level varies according to  $P(t)$ . If  $r_e(t)$  is defined analogously, the price of the bond selling for  $1/r_e(0)$  at time 0 will be

$$\frac{P(t)}{P(0) r_e(t)}$$

at time  $t$ , where the ratio of prices is required to adjust for any change in the price level. The nominal stream purchased for \$1.00 at time zero then consists of

$$r_e(0) \cdot \frac{P(t)}{P(0)} + \frac{r_e(0)}{P(0)} \cdot d \left[ \frac{P(t)}{r_e(t)} \right] = r_e(0) \cdot \frac{P(t)}{P(0)} + \frac{r_e(0)}{r_e(t)} \cdot \frac{1}{P(0)} \cdot \frac{dP(t)}{dt} - \frac{r_e(0)}{P(0)} \cdot \frac{r_e(0)}{r_e^2(t)} \cdot \frac{dr_e(t)}{dt}. \quad (4)$$

Once again we can approximate this functional by its value at time zero, which is

$$r_e + \frac{1}{P} \frac{dP}{dt} - \frac{1}{r_e} \frac{dr_e}{dt}. \quad (5)$$

This sum, together with  $P$  already introduced, defines the "real" return from holding \$1.00 of wealth in the form of equities.

(iv) Physical goods held by ultimate wealth-owning units are similar to equities except that the annual stream they yield is in kind rather than in money. In terms of nominal units, this return, like that from equities, depends on the behavior of prices. In addition, like equities, physical goods must be regarded as yielding a nominal return in the form of appreciation or depreciation in money value. If we suppose the price level  $P$ , introduced earlier, to apply equally to the value of these physical goods, then, at time zero,

$$\frac{1}{P} \frac{dP}{dt} \quad (6)$$

is the size of this nominal return per \$1.00 of physical goods.<sup>2</sup> Together with  $P$ , it defines the "real" return from holding \$1.00 in the form of physical goods.

(v) Since there is only a limited market in human capital, at least in modern non-slave societies, we cannot very well define in market prices the terms of substitution of human capital for other forms of capital and so cannot define at any time the physical unit of capital corresponding to \$1.00 of human capital. There are some possibilities of substituting non-human capital for human capital in an individual's wealth holdings, as, for example, when he enters into a contract to render personal services for a specified period in return for a definitely specified number of periodic payments, the number not depending on his being physically capable of rendering the services. But, in the main, shifts between human capital and other forms must take place through direct investment and disinvestment in the human agent, and we may as well treat this as if it were the only way. With respect to this form of capital, therefore, the restriction or obstacles affecting the alternative compositions of wealth available to the individual cannot be expressed in terms of market prices or rates of return. At any one point in time there is some division between human and non-human wealth in his portfolio of assets; he may be able to change this over time, but we shall treat it as given at a point in time. Let  $w$  be the ratio of non-human to human wealth or, equivalently, of income from non-human wealth to income from human wealth, which means that it is closely allied to what is usually defined as the ratio of wealth to income. This is, then, the variable that needs to be taken into account so far as human wealth is concerned.

6. The tastes and preferences of wealth-owning units for the service streams arising from different forms of wealth must in general simply be taken for granted as determining the form of the demand function. In order to give the theory empirical content, it will generally have to be supposed that tastes are constant over significant stretches of space and time. However, explicit allowance can be made for some changes in tastes in so far as such changes are linked with objective circumstances. For example, it seems reasonable that, other things the same, individuals want

2. In principle, it might be better to let  $P$  refer solely to the value of the services of physical goods, which is essentially what it refers to in the preceding cases, and to allow for the fact that the prices of the capital goods themselves must vary also with the rate of capitalization, so that the prices of services and their sources vary at the same rate only if the relevant interest rate is constant. I have neglected this refinement for simplicity; the neglect can perhaps be justified by the rapid depreciation of many of the physical goods held by final wealth-owning units.

to hold a larger fraction of their wealth in the form of money when they are moving around geographically or are subject to unusual uncertainty than otherwise. This is probably one of the major factors explaining a frequent tendency for money holdings to rise relative to income during wartime. But the extent of geographic movement, and perhaps of other kinds of uncertainty, can be represented by objective indexes, such as indexes of migration, miles of railroad travel, and the like. Let  $u$  stand for any such variables that can be expected to affect tastes and preferences (for "utility" determining variables).

7. Combining 4, 5, and 6 along the lines suggested by 3 yields the following demand function for money:

$$M = f \left( P, r_b - \frac{1}{r_b} \frac{dr_b}{dt}, r_s + \frac{1}{P} \frac{dP}{dt} - \frac{1}{r_s} \frac{dr_s}{dt}, \frac{1}{P} \frac{dP}{dt}; w; \frac{V}{P}; u \right). \quad (7)$$

A number of observations are in order about this function.

(i) Even if we suppose prices and rates of interest unchanged, the function contains three rates of interest: two for specific types of assets,  $r_b$  and  $r_s$ , and one intended to apply to all types of assets,  $r$ . This general rate,  $r$ , is to be interpreted as something of a weighted average of the two special rates plus the rates applicable to human wealth and to physical goods. Since the latter two cannot be observed directly, it is perhaps best to regard them as varying in some systematic way with  $r_b$  and  $r_s$ . On this assumption, we can drop  $r$  as an additional explicit variable, treating its influence as fully taken into account by the inclusion of  $r_b$  and  $r_s$ .

(ii) If there were no differences of opinion about price movements and interest-rate movements, and bonds and equities were equivalent except that the former are expressed in nominal units, arbitrage would of course make

$$r_b - \frac{1}{r_b} \frac{dr_b}{dt} = r_s + \frac{1}{P} \frac{dP}{dt} - \frac{1}{r_s} \frac{dr_s}{dt}, \quad (8)$$

or, if we suppose rates of interest either stable or changing at the same percentage rate,

$$r_b = r_s + \frac{1}{P} \frac{dP}{dt}. \quad (9)$$

that is, the "money" interest rate equal to the "real" rate plus the percentage rate of change of prices. In application the rate of change of prices must be interpreted as an "expected" rate of change and differences of opinion cannot be neglected, so we cannot suppose (9) to hold; indeed,

one of the most consistent features of inflation seems to be that it does not.

(iii) If the range of assets were to be widened to include promises to pay specified sums for a finite number of time units—"short-term" securities as well as "consols"—the rates of change of  $r_b$  and  $r_s$  would be reflected in the difference between long and short rates of interest. Since at some stage it will doubtless be desirable to introduce securities of different time duration (see point 23 below), we may simplify the present exposition by restricting it to the case in which  $r_b$  and  $r_s$  are taken to be stable over time. Since the rate of change in prices is required separately in any event, this means that we can replace the cumbrous variables introduced to designate the nominal return on bonds and equities simply by  $r_b$  and  $r_s$ .

(iv)  $Y$  can be interpreted as including the return to all forms of wealth, including money and physical capital goods owned and held directly by ultimate wealth-owning units, and so  $Y/r$  can be interpreted as an estimate of total wealth, only if  $Y$  is regarded as including some imputed income from the stock of money and directly owned physical capital goods. For monetary analysis the simplest procedure is perhaps to regard  $Y$  as referring to the return to all forms of wealth other than the money held directly by ultimate wealth-owning units, and so  $Y/r$  as referring to total remaining wealth.

8. A more fundamental point is that, as in all demand analyses resting on maximization of a utility function defined in terms of "real" magnitudes, this demand equation must be considered independent in any essential way of the nominal units used to measure money variables. If the unit in which prices and money income are expressed is changed, the amount of money demanded should change proportionately. More technically, equation (7) must be regarded as homogeneous of the first degree in  $P$  and  $Y$ , so that

$$\begin{aligned} f(\lambda P, r_b, r_s, \frac{1}{P} \frac{dP}{dt}; w; \lambda Y; u) \\ = \lambda f(P, r_b, r_s, \frac{1}{P} \frac{dP}{dt}; w; Y; u). \end{aligned} \quad (10)$$

where the variables within the parentheses have been rewritten in simpler form in accordance with comments 7 (i) and 7 (iii).

3. See Reuben Kessel, "Inflation: Theory of Wealth Distribution and Application in Private Investment Policy" (unpublished doctoral dissertation, University of Chicago).

This characteristic of the function enables us to rewrite it in two alternative and more familiar ways.

(i) Let  $\lambda = 1/P$ . Equation (7) can then be written

$$\frac{M}{P} = f\left(r_b, r_s, \frac{1}{P} \frac{dP}{dt}; w; \frac{Y}{P}; u\right). \quad (11)$$

In this form the equation expresses the demand for real balances as a function of "real" variables independent of nominal monetary values.

(ii) Let  $\lambda = 1/Y$ . Equation (7) can then be written

$$\begin{aligned} \frac{M}{Y} = f\left(r_b, r_s, \frac{1}{P} \frac{dP}{dt}, w, \frac{P}{Y}, u\right) \\ = \frac{1}{v\left(r_b, r_s, \frac{1}{P} \frac{dP}{dt}, w, \frac{P}{Y}, u\right)}, \end{aligned} \quad (12)$$

or

$$Y = v\left(r_b, r_s, \frac{1}{P} \frac{dP}{dt}, w, \frac{P}{Y}, u\right) \cdot M. \quad (13)$$

In this form the equation is in the usual quantity theory form, where  $v$  is income velocity.

9. These equations are, to this point, solely for money held directly by ultimate wealth-owning units. As noted, money is also held by business enterprises as a productive resource. The counterpart to this business asset in the balance sheet of an ultimate wealth-owning unit is a claim other than money. For example, an individual may buy bonds from a corporation, and the corporation use the proceeds to finance the money holdings which it needs for its operations. Of course, the usual difficulties of separating the accounts of the business and its owner arise with unincorporated enterprises.

10. The amount of money that it pays business enterprises to hold depends, as for any other source of productive services, on the cost of the productive services, the cost of substitute productive services, and the value product yielded by the productive service. Per dollar of money held, the cost depends on how the corresponding capital is raised—whether by raising additional capital in the form of bonds or equities, by substituting cash for real capital goods, etc. These ways of financing money holdings are much the same as the alternative forms in which the ultimate wealth-owning unit can hold its non-human wealth, so that the variables  $r_b$ ,  $r_s$ ,  $P$ , and  $(1/P)(dP/dt)$  introduced into (7) can be taken to represent the cost to the business enterprise of holding money. For some purposes, however, it may be desirable to distinguish between the rate of return re-

ceived by the lender and the rate paid by the borrower; in which case it would be necessary to introduce an additional set of variables.

Substitutes for money as a productive service are numerous and varied, including all ways of economizing on money holdings by using other resources to synchronize more closely payments and receipts, reduce payment periods, extend use of book credit, establish clearing arrangements, and so on in infinite variety. There seem no particularly close substitutes whose prices deserve to be singled out for inclusion in the business demand for money.

The value product yielded by the productive services of money per unit of output depends on production conditions: the production function. It is likely to be especially dependent on features of production conditions affecting the smoothness and regularity of operations as well as on those determining the size and scope of enterprises, degree of vertical integration, etc. Again there seem no variables that deserve to be singled out on the present level of abstraction for special attention; these factors can be taken into account by interpreting  $u$  as including variables affecting not only the tastes of wealth-owners but also the relevant technological conditions of production. Given the amount of money demanded per unit of output, the total amount demanded is proportional to total output, which can be represented by  $Y$ .

11. One variable that has traditionally been singled out in considering the demand for money on the part of business enterprises is the volume of transactions, or of transactions per dollar of final products; and, of course, emphasis on transactions has been carried over to the ultimate wealth-owning unit as well as to the business enterprise. The idea that renders this approach attractive is that there is a mechanical link between a dollar of payments per unit time and the average stock of money required to effect it—a fixed technical coefficient of production, as it were. It is clear that this mechanical approach is very different in spirit from the one we have been following. On our approach, the average amount of money held per dollar of transactions is itself to be regarded as a resultant of an economic equilibrating process, not as a physical datum. If, for whatever reason, it becomes more expensive to hold money, then it is worth devoting resources to effecting money transactions in less expensive ways or to reducing the volume of transactions per dollar of final output. In consequence, our ultimate demand function for money in its most general form does not contain as a variable the volume of transactions or of transactions per dollar of final output; it contains rather those more basic technical and cost conditions that affect the costs of conserving money, be it by changing the average amount of money held per dollar of transac-

tions per unit time or by changing the number of dollars of transactions per dollar of final output. This does not, of course, exclude the possibility that, for a particular problem, it may be useful to regard the transactions variables as given and not to dig beneath them and so to include the volume of transactions per dollar of final output as an explicit variable in a special variant of the demand function.

Similar remarks are relevant to various features of payment conditions, frequently described as "institutional conditions," affecting the velocity of circulation of money and taken as somehow mechanically determined—such items as whether workers are paid by the day, or week, or month; the use of book credit; and so on. On our approach these, too, are to be regarded as resultants of an economic equilibrating process, not as physical data. Lengthening the pay period, for example, may save book-keeping and other costs to the employer, who is therefore willing to pay somewhat more than in proportion for a longer than a shorter pay period; on the other hand, it imposes on employees the cost of holding larger cash balances or providing substitutes for cash, and they therefore want to be paid more than in proportion for a longer pay period. Where these will balance depends on how costs vary with length of pay period. The cost to the employee depends in considerable part on the factors entering into his demand curve for money for a fixed pay period. If he would in any event be holding relatively large average balances, the additional costs imposed by a lengthened pay period tend to be less than if he would be holding relatively small average balances, and so it will take less of an inducement to get him to accept a longer pay period. For given cost savings to the employer, therefore, the pay period can be expected to be longer in the first case than in the second. Surely, the increase in the average cash balance over the past century in this country that has occurred for other reasons has been a factor producing a lengthening of pay periods and not the other way around. Or, again, experience in hyperinflation shows how rapidly payment practices change under the impact of drastic changes in the cost of holding money.

12. The upshot of these considerations is that the demand for money on the part of business enterprises can be regarded as expressed by a function of the same kind as equation (7), with the same variables on the right-hand side. And, like (7), since the analysis is based on informed maximization of returns by enterprises, only "real" quantities matter, so it must be homogeneous of the first degree in  $Y$  and  $P$ . In consequence, we can interpret (7) and its variants (11) and (13) as describing the demand for money on the part of a business enterprise as well as on the part of an

ultimate wealth-owning unit, provided only that we broaden our interpretation of  $u$ .

13. Strictly speaking, the equations (7), (11), and (13) are for an individual wealth-owning unit or business enterprise. If we aggregate (7) for all wealth-owning units and business enterprises in the society, the result, in principle, depends on the distribution of the units by the several variables. This raises no serious problem about  $P$ ,  $r$ , and  $r_u$  for these can be taken as the same for all, or about  $u$ , for this is an unspecified portmanteau variable to be filled in as the occasion demands. We have been interpreting  $(1/P)(dP/dt)$  as the expected rate of price rise, so there is no reason why this variable should be the same for all, and  $w$  and  $Y$  clearly differ substantially among units. An approximation is to neglect these difficulties and take (7) and the associated (11) and (13) as applying to the aggregate demand for money, with  $(1/P)(dP/dt)$  interpreted as some kind of an average expected rate of change of prices,  $w$  as the ratio of total income from non-human wealth to income from human wealth, and  $Y$  as aggregate income. This is the procedure that has generally been followed, and it seems the right one until serious departures between this linear approximation and experience make it necessary to introduce measures of dispersion with respect to one or more of the variables.

14. It is perhaps worth noting explicitly that the model does not use the distinction between "active balances" and "idle balances" or the closely allied distinction between "transaction balances" and "speculative balances" that is so widely used in the literature. The distinction between money holdings of ultimate wealth-owners and of business enterprises is related to this distinction but only distantly so. Each of these categories of money-holders can be said to demand money partly from "transaction" motives, partly from "speculative" or "asset" motives, but dollars of money are not distinguished according as they are said to be held for one or the other purpose. Rather, each dollar is, as it were, regarded as rendering a variety of services, and the holder of money as altering his money holdings until the value to him of the addition to the total flow of services produced by adding a dollar to his money stock is equal to the reduction in the flow of services produced by subtracting a dollar from each of the other forms in which he holds assets.

15. Nothing has been said above about "banks" or producers of money. This is because their main role is in connection with the supply of money rather than the demand for it. Their introduction does, however, blur some of the points in the above analysis: the existence of banks enables productive enterprises to acquire money balances without raising capital from ultimate wealth-owners. Instead of selling claims (bonds or

## *The Quantity Theory of Money—A Restatement*

equities) to them, it can sell its claims to banks, getting "money" in exchange: in the phrase that was once so common in textbooks on money, the bank coins specific liabilities into generally acceptable liabilities. But this possibility does not alter the preceding analysis in any essential way.

16. Suppose the supply of money in nominal units is regarded as fixed or more generally autonomously determined. Equation (13) then defines the conditions under which this nominal stock of money will be the amount demanded. Even under these conditions, equation (13) alone is not sufficient to determine money income. In order to have a complete model for the determination of money income, it would be necessary to specify the determinants of the structure of interest rates, of real income, and of the path of adjustment in the price level. Even if we suppose interest rates determined independently—by productivity, thrift, and the like—and real income as also given by other forces, equation (13) only determines a unique equilibrium level of money income if we mean by this the level at which prices are stable. More generally, it determines a time path of money income for given initial values of money income.

In order to convert equation (13) into a "complete" model of income determination, therefore, it is necessary to suppose either that the demand for money is highly inelastic with respect to the variables in  $v$  or that all these variables are to be taken as rigid and fixed.

17. Even under the most favorable conditions, for example, that the demand for money is quite inelastic with respect to the variables in  $v$ , equation (13) gives at most a theory of money income: it then says that changes in money income mirror changes in the nominal quantity of money. But it tells nothing about how much of any change in  $Y$  is reflected in real output and how much in prices. To infer this requires bringing in outside information, as, for example, that real output is at its feasible maximum, in which case any increase in money would produce the same or a larger percentage increase in prices; and so on.

18. In light of the preceding exposition, the question arises what it means to say that someone is or is not a "quantity theorist." Almost every economist will accept the general lines of the preceding analysis on a purely formal and abstract level, although each would doubtless choose to express it differently in detail. Yet there clearly are deep and fundamental differences about the importance of this analysis for the understanding of short- and long-term movements in general economic activity. This difference of opinion arises with respect to three different issues: (i) the stability and importance of the demand function for money; (ii) the independence of the factors affecting demand and supply; and (iii) the form of the demand function or related functions.



(i) The quantity theorist accepts the empirical hypothesis that the demand for money is highly stable—more stable than functions such as the consumption function that are offered as alternative key relations. This hypothesis needs to be hedged on both sides. On the one side, the quantity theorist need not, and generally does not, mean that the real quantity of money demanded per unit of output, or the velocity of circulation of money, is to be regarded as numerically constant over time; he does not, for example, regard it as a contradiction to the stability of the demand for money that the velocity of circulation of money rises drastically during hyperinflation. For the stability he expects is in the functional relation between the quantity of money demanded and the variables that determine it, and the sharp rise in the velocity of circulation of money during hyperinflation is entirely consistent with a stable functional relation, as Cagan so clearly demonstrates in his essay. On the other side, the quantity theorist must sharply limit, and be prepared to specify explicitly, the variables that it is empirically important to include in the function. For to expand the number of variables regarded as significant is to empty the hypothesis of its empirical content; there is indeed little if any difference between asserting that the demand for money is highly unstable and asserting that it is a perfectly stable function of an indefinitely large number of variables.

The quantity theorist not only regards the demand function for money as stable; he also regards it as playing a vital role in determining variables that he regards as of great importance for the analysis of the economy as a whole, such as the level of money income or of prices. It is this that leads him to put greater emphasis on the demand for money than on, let us say, the demand for pins, even though the latter might be as stable as the former. It is not easy to state this point precisely, and I cannot pretend to have done so. (See item [iii] below for an example of an argument against the quantity theorist along these lines.)

The reaction against the quantity theory in the 1930's came largely, I believe, under this head. The demand for money, it was asserted, is a will-o'-the-wisp, shifting erratically and unpredictably with every rumor and expectation; one cannot, it was asserted, reliably specify a limited number of variables on which it depends. However, although the reaction came under this head, it was largely rationalized under the two succeeding heads.

(ii) The quantity theorist also holds that there are important factors affecting the supply of money that do not affect the demand for money. Under some circumstances these are technical conditions affecting the supply of specie; under others, political or psychological conditions determining the policies of monetary authorities and the banking system. A stable

demand function is useful precisely in order to trace out the effects of changes in supply, which means that it is useful only if supply is affected by at least some factors other than those regarded as affecting demand.

The classical version of the objection under this head to the quantity theory is the so-called real-bills doctrine: that changes in the demand for money call forth corresponding changes in supply and that supply cannot change otherwise, or at least cannot do so under specified institutional arrangements. The forms which this argument takes are legion and are still widespread. Another version is the argument that the "quantity theory" cannot "explain" large price rises, because the price rise produces both the increase in demand for nominal money holdings and the increase in supply of money to meet it; that is, implicitly that the same force affects both the demand for and the supply of money, and in the same way.

(iii) The attack on the quantity theory associated with the Keynesian underemployment analysis is based primarily on an assertion about the form of (7) or (11). The demand for money, it is said, is infinitely elastic at a "small" positive interest rate. At this interest rate, which can be expected to prevail under underemployment conditions, changes in the real supply of money, whether produced by changes in prices or in the nominal stock of money, have no effect on anything. This is the famous "liquidity trap." A rather more complex version involves the shape of other functions as well: the magnitudes in (7) other than "the" interest rate, it is argued, enter into other relations in the economic system and can be regarded as determined there; the interest rate does not enter into these other functions; it can therefore be regarded as determined by this equation. So the only role of the stock of money and the demand for money is to determine the interest rate.

19. The proof of this pudding is in the eating; and the essays in this book contain much relevant food, of which I may perhaps mention three particularly juicy items.

One cannot read Lerner's description of the effects of monetary reform in the Confederacy in 1864 without recognizing that at least on occasion the supply of money can be a largely autonomous factor and the demand for money highly stable even under extraordinarily unstable circumstances. After three years of war, after widespread destruction and military reverses, in the face of impending defeat, a monetary reform that succeeded in reducing the stock of money halted and reversed for some months a rise in prices that had been going on at the rate of 11 per cent a month most of the war! It would be hard to construct a better controlled experiment to demonstrate the critical importance of the supply of money.

On the other hand, Klein's examination of German experience in World

that  $Y$  was introduced into equation (7) as an index of wealth. This has important implications for the measure or concept of income that is relevant. What is required by the theoretical analysis is not usual measured income—which in the main corresponds to current receipts corrected for double counting—but a longer term concept, "expected income," or what I have elsewhere called "permanent income."<sup>4</sup> Now suppose that the variables in the  $v$  function of (13) are unchanged for a period. The ratio of  $Y$  to  $M$  would then be unchanged, provided  $Y$  is *permanent* income. Velocity as Selden computes it is the ratio of *measured* income to the stock of money and would not be unchanged. When measured income was above permanent income, measured velocity would be relatively high, and conversely. Now measured income is presumably above permanent income at cyclical peaks and below permanent income at cyclical troughs. The observed positive conformity of measured velocity to cyclical changes of income may therefore reflect simply the difference between measured income and the concept relevant to equation (13).

that  $Y$  was introduced into equation (7) as an index of wealth. This has important implications for the measure or concept of income that is relevant. What is required by the theoretical analysis is not usual measured income—which in the main corresponds to current receipts corrected for double counting—but a longer term concept, "expected income," or what I have elsewhere called "permanent income."<sup>4</sup> Now suppose that the variables in the  $v$  function of (13) are unchanged for a period. The ratio of  $Y$  to  $M$  would then be unchanged, provided  $Y$  is *permanent* income. Velocity as Selden computes it is the ratio of *measured* income to the stock of money and would not be unchanged. When measured income was above permanent income, measured velocity would be relatively high, and conversely. Now measured income is presumably above permanent income at cyclical peaks and below permanent income at cyclical troughs. The observed positive conformity of measured velocity to cyclical changes of income may therefore reflect simply the difference between measured income and the concept relevant to equation (13).

23. Another point that is raised by Selden's work is the appropriate division of wealth into forms of assets. The division suggested above is, of course, only suggestive. Selden finds more useful the distinction between "short-term" and "long-term" bonds; he treats the former as "substitutes for money" and calls the return on the latter "the cost of holding money." He finds both to be significantly related to the quantity of money demanded. It was suggested above that this is also a way to take into account expectations about changes in interest rates. Similarly, there is no hard-and-fast line between "money" and other assets, and for some purposes it may be desirable to distinguish between different forms of "money" (e.g., between currency and deposits). Some of these forms of money may pay interest or may involve service charges, in which case the positive or negative return will be a relevant variable in determining the division of money holdings among various forms.

24. By concentrating on hyperinflations, Cagan was able to bring into sharp relief a variable whose effect is generally hard to evaluate, namely, the rate of change of prices. The other side of this coin is the necessity of neglecting practically all the remaining variables. His device for estimating expected rates of change of prices from actual rates of change, which works so well for his data, can be carried over to other variables as well and so is likely to be important in fields other than money. I have already used it to estimate "expected income" as a determinant of consumption.<sup>5</sup>

25. Selden finds that for cyclical periods velocity rises during expansions and falls during contractions, a result that at first glance seems to contradict the secular result just cited. However, there is an alternative explanation entirely consistent with the secular result. It will be recalled

4. See Milton Friedman, *A Theory of the Consumption Function*, forthcoming publication of the Princeton University Press for the National Bureau of Economic Research.

5. See *ibid.*

series in a demand function for money along the lines suggested above (in point 22).

Cagan's results make it clear that changes in the rate of change of prices, or in the return to an alternative form of holding wealth, have the expected effect on the quantity of money demanded: the higher the rate of change of prices, and thus the more attractive the alternative, the less the quantity of money demanded. This result is important not only directly but also because it is indirectly relevant to the effect of changes in the returns to other alternatives, such as rates of interest on various kinds of bonds. Our evidence on these is in some way less satisfactory because they have varied over so much smaller a range; tentative findings that the effect of changes in them is in the expected direction are greatly strengthened by Cagan's results.

One point which is suggested by the inapplicability of Cagan's relations to the final stages of the hyperinflation is that it may at times be undesirable to replace the whole expected pattern of price movements by the rate of change expected at the moment, as Cagan does and as is done in point 5 above. For example, a given rate of price rise, expected to continue, say, for only a day, and to be followed by price stability, will clearly mean a higher (real) demand for money than the same rate of price rise expected to continue indefinitely; it will be worth incurring greater costs to avoid paying the latter than the former price. This is the same complication as occurs in demand analysis for a consumer good when it is necessary to include not only the present price but also past prices or future expected prices. This point may help explain not only Cagan's findings for the terminal stages but also Selken's findings that the inclusion of the rate of change of prices as part of the cost of holding money worsened rather than improved his estimated relations, though it may be that this result arises from a different source, namely, that it takes substantial actual rates of price change to produce firm enough and uniform enough expectations about price behavior for this variable to play a crucial role.

Similar comments are clearly relevant for expected changes in interest rates.

25. One of the chief reproaches directed at economics as an allegedly empirical science is that it can offer so few numerical "constants," that it has isolated so few fundamental regularities. The field of money is the chief example one can offer in rebuttal: there is perhaps no other empirical relation in economics that has been observed to recur so uniformly under so wide a variety of circumstances as the relation between substantial

changes over short periods in the stock of money and in prices; the one is invariably linked with the other and is in the same direction; this uniformity is, I suspect, of the same order as many of the uniformities that form the basis of the physical sciences. And the uniformity is in more than one direction. There is an extraordinary empirical stability and regularity to such magnitudes as income velocity that cannot but impress anyone who works extensively with monetary data. This very stability and regularity contributed to the downfall of the quantity theory, for it was overstated and expressed in unduly simple form; the numerical value of the velocity itself, whether income or transactions, was treated as a natural "constant." Now this it is not; and its failure to be so, first during and after World War I and then, to a lesser extent, after the crash of 1929, helped greatly to foster the reaction against the quantity theory. The studies in this volume are premised on a stability and regularity in monetary relations of a more sophisticated form than a numerically constant velocity. And they make, I believe, an important contribution toward extracting this stability and regularity, toward isolating the numerical "constants" of monetary behavior. It is by this criterion at any rate that I, and I believe also their authors, would wish them to be judged.

I began this Introduction by referring to the tradition in the field of money at Chicago and to the role of faculty members in promoting it. I think it is fitting to end the Introduction by emphasizing the part which students have played in keeping that tradition alive and vigorous. The essays that follow are one manifestation. Unpublished doctoral dissertations on money are another. In addition, I wish especially to express my own personal appreciation to the students who have participated with me in the Workshop in Money and Banking, of which this volume is the first published fruit. I owe a special debt to David I. Fand, Phillip Cagan, Gary Becker, David Meischman, and Raymond Zelder, who have at various times helped me to conduct it.

We all of us are indebted also to the Rockefeller Foundation for financial assistance to the Workshop in Money and Banking. This assistance helped to finance some of the research reported in this book and has made possible its publication.